Area-Based Models of Highway Growth

David Levinson[1] and Wei Chen[2]

Abstract: Empirical data and statistical models are employed to predict where new highway routes are most likely to be located. The land use, population distribution, and highway network for the Twin Cities' Metro Area from 1958 to 1990 are used. Binary logit models estimate the likelihood a particular cell will see the construction of divided highways and secondary highways. The results show that the area's land-use attributes and population density levels do significantly affect the likelihood of adding new highway routes.

DOI: 10.1061/(ASCE)0733-9488(2007)133:4(250)

CE Database subject headings: Land usage; Geographic information systems; Highways; Routing; Statistics; Predictions.

Introduction

Determining where new rights of way will be located or which existing routes will be widened is a problem of significant importance to any effort to understand, forecast, and plan the growth of cities. However, models of network growth are few in number, most are theoretical or conceptual (e.g., Taaffe et al. 1963; Helbing et al. 1997; Yamins et al. 2003; Yerra and Levinson 2005; Levinson and Yerra 2006), while only a few are empirical (Garrison and Marble 1962, 1965; Levinson and Karamalaputi 2003a,b). As a result, there is little guidance for researchers or professionals about the significant variables or the magnitude of relations between those variables that explain network investments. This study employs available empirical data and statistical models to test the factors that explain where new highway routes are most likely to be built. Binary logit models estimate the likelihood that divided highways and secondary highways will be constructed in particular small geographic cells based on land use, population distribution, and highway network data. This analysis is intended to test particular hypotheses regarding the effect of land use and population on network investment, rather than form a general model of network growth. In the following sections, the data for modeling divided highways and secondary highways are described, and the hypotheses, models, and results are presented sequentially.

Description of Data Sets

High-quality geographic information system (GIS) maps were digitized from paper maps for this study (Twin Cities Metropolitan Council 1958, 1968, 1978; Minnesota Department of Transortation 1962, 1965, 1968, 1971, 1975, 1978, 1981, 1985, 1990; US Census Bureau 1960, 1970, 1980). The study period begins with 1958 when the earliest land-use map was created for the Twin Cities Metro Area. Then a lattice layer composed of 30,729 square cells ($0.141\ \mathrm{km}^2$) ($375\ \mathrm{m/side}$) is created. Each of the population distribution layers, land-use layers, and highway network layers are merged into the lattice layer, so that each cell of the lattice layer contains the spatial information of population, land use, and highways.

Each cell can be viewed as one observation. Land use, population distribution, and highway network data from the base years and highway network data from the test years will be used in the models to estimate the probability that new highway growth will occur in a cell. Table 1 summarizes the base year and test year data.

Divided Highways

Hypotheses

Hypotheses of the growth tendency of divided highways are presented as follows.

Agglomeration. Agglomeration is the phenomenon of roads of a particular class being built near (or connecting to) similar roads. The agglomeration of divided highways includes both the emergence of alternative routes and also the extension of the existing corridors. First, as major commuting corridors, divided highways typically locate near regions with relatively intense economic activities. Moreover, divided highways may lead to further economic development nearby, which means more traffic demand in the neighborhood of existing corridors. Second, connectivity induces the further extension of existing corridors or the addition of new links adjoining the old ones. After new routes appear, the boosted economic activity and traffic growth nearby may lead to another round of agglomeration growth.

Generally we expect that cells neighboring existing divided highways are more likely to exhibit network growth. To test the hypothesis of agglomeration, a 0.5-km buffer area of the existing divided and undivided highways is made. (The reason for includ-

[1]Dept. of Civil Engineering, Univ. of Minnesota, 500 Pillsbury Dr. SE, Minneapolis, MN 55455 (corresponding author). E-mail: dlevinson@umn.edu

[2]SEMCOG, 535 Griswold St., Suite 300, Detroit, MI 48226.

Note. Discussion open until May 1, 2008. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this technical note was submitted for review and possible publication on March 7, 2006; approved on November 13, 2006. This technical note is part of the ***Journal of Urban Planning and Development***, Vol. 133, No. 4, December 1, 2007. ©ASCE, ISSN 0733-9488/2007/4-250–254/$25.00.

Table 1. Base Years' Data and Predicted Years' Data Used for Model Estimation

Base years' land-use, population distribution, and highway network data	Predicted years' highway network data
(a) 3-year new highway growth probability estimation	
1968 Land use	1971 Highway network
1970 Population distribution	
1968 Highway network	
1978 Land use	1981 Highway network
1980 Population distribution	
1978 Highway network	
(b) 5-year new highway growth probability estimation	
1958 Land use	1965 Highway network
1960 Population distribution	
1962 Highway network[a]	
1968 Land use	1975 Highway network
1970 Population distribution	
1968 Highway network	
1978 Land use	1985 Highway network
1980 Population distribution	
1978 Highway network	
(c) 10-year new highway growth probability estimation	
1958 Land use	1968 Highway network
1960 Population distribution	
1962 Highway network	
1968 Land use	1978 Highway network
1970 Population distribution	
1968 Highway network	
1978 Land use	1990 Highway network
1980 Population distribution	
1978 Highway network	

[a]1958 Highway Network should be used as the base year data, but before 1962, the Twin Cites Metro Area just covered very small area in the Minnesota Official Highway Maps and more than half of the observations would be dropped if 1958 Highway Network were used. Therefore, 1962 Highway Network is used as a substitute.

ing undivided highways is that undivided highways may be upgraded to divided highways.) Variable $A=1$ if the observed cell is within or intersects the buffer area, otherwise $A=0$. The cells with $A=1$ should be associated with higher growth than the cells with $A=0$.

Population Density. Cells are classified into four area types, P_S [sparsely populated (0–50 people/cell)]; P_L [(low population (51–100 people)]; P_M [medium population (101–300 people)]; and P_H [high population (more than 300 people)]. We expect networks are more likely to expand in low and medium populated areas than sparsely or highly populated areas. For high-density areas, although their neighboring areas may have more traffic demand, they are usually associated with high land prices and costly relocation. Therefore, the final decision of route development in high-density areas should balance demand, costs, and the availability of cheaper alternative routes.

Employment Zones. As mentioned before, divided highways mainly serve commuting traffic. Therefore, we expect the further growth of divided highways tends to be close to employment zones. To test this hypothesis, variable $U_E=1$ if the observed cell is within or intersects employment zones, otherwise $U_E=0$. The cells with $U_E=1$ should have a higher probability of divided highway growth than the cells with $U_E=0$.

Commercial Zones. We expect that commercial zones and their neighborhood are associated with more network growth. To test this hypothesis, variable $U_C=1$ if the observed cell is within or intersects commercial zones, otherwise $U_C=0$. The cells with $U_C=1$ should have a higher probability of divided highway growth than the cells with $U_C=0$.

Agricultural Areas. Variable U_A is defined as the share of agricultural areas within each cell, and we expect that agricultural areas are negatively associated with route growth. However, it should be noted that although agricultural areas typically have low traffic demand, they are also the areas of low land prices; furthermore, one purpose of divided highways is to connect urban and suburban areas and spur economic development of the undeveloped areas. These factors may lead to route growth in the agricultural areas.

Water Areas. Water is a barrier for highway development. Variable U_W is defined as the percentage of water area within each cell, and we expect that U_W is negatively related to route growth.

Model

A binary logit model estimates the divided highway growth based on the population distribution, land use, and highway network data of the base years. It should be noted that the model estimates the probability of divided highway growth in each cell, but it does not estimate the extent of growth. The extent of growth is influenced by many factors (such as the direction of the highway segment, the connection with other links, the geographical or geological conditions, etc.) that cannot be controlled in this study but should be addressed in future research.

To diagnose potential multicollinearity, we examined the correlations among the variables and none of them was larger than 0.60. We did not find the symptoms of multicollinearity [such as inflated standard errors, excessive logit iterations (more than 10 or 15 times) or unreasonable statistical results of the critical independent variable(s), etc.]. So multicollinearity should not cause disturbance in this case.

Interaction terms have been tested. Since the addition of the interaction terms did not significantly improve the overall model performance and most of the interaction terms are insignificant, the interaction terms are not included in the final model. Of the models tested, the following model is the best in overall fit, and the regression results of the model are presented in Table 2

$$G_D = f(P_S,P_L,P_M,P_H,A,U_E,U_C,U_A,U_W,L_I,L_D,L_U,L_S,D,Y)$$

where G_D (dependent variable)=divided highway growth, if from the base year to the test year there is growth in divided highways in the observed cell, $G_D=1$, otherwise $G_D=0$; P_S, P_L, P_M, and P_H=population—all the cells are classified into four groups, P_S (sparsely populated area), P_L (low population area), P_M (medium population area), and P_H (high population area); A=agglomeration—if the observed cell is within or intersects the 0.5-km buffer area of the divided and undivided highways of the base year, $A=1$, otherwise $A=0$; U_E, U_C, U_A, and U_W=land use—they are defined as follows: U_E=if the observed cell is within or intersects employment zones (including airports) of the base year, $U_E=1$, otherwise $U_E=0$; U_C=if the observed cell is within or intersects commercial areas of the base year, $U_C=1$, otherwise

Table 2. Logit Regression Results for Divided Highway Growth

Dependent variable $=G_D$		Logit regression results								
		3_year_growth			5_year_growth			10_year_growth		
Independent variable		Odds ratio	Coef.	$P>\|z\|$	Odds ratio	Coef.	$P>\|z\|$	Odds ratio	Coef.	$P>\|z\|$
Low population area	P_L	1.945	0.665[a]	0.000	2.192	0.785[a]	0.000	2.013	0.699[a]	0.000
Medium population area	P_M	1.565	0.448[a]	0.004	1.868	0.625[a]	0.000	1.579	0.457[a]	0.000
High population area	P_H	0.595	−0.519[a]	0.017	0.529	−0.636[a]	0.000	0.408	−0.897[a]	0.000
Agglomeration Predictor	A	2.618	0.962[a]	0.000	1.000	0.000[a]	0.019	1.000	0.000[a]	0.000
Employment zones [1,0]	U_E	1.213	0.193[a]	0.069	1.068	0.066	0.435	1.112	0.106	0.161
Commercial areas [1,0] [[1,0][1,0]	U_C	1.711	0.537[a]	0.000	1.435	0.361[a]	0.000	1.288	0.253[a]	0.004
% Agriculture area	U_A	0.845	−0.168	0.310	1.005	0.005	0.965	0.943	−0.059	0.560
% Water area	U_W	0.204	−1.590[a]	0.004	0.152	−1.887[a]	0.000	0.134	−2.010[a]	0.000
Km Interstates	L_I	1.000	0.000	0.988	1.000	0.000	0.532	1.000	0.000	0.650
Km Divided highways	L_D	1.001	0.001[a]	0.002	1.002	0.002[a]	0.000	1.002	0.002[a]	0.000
Km Undivided highways	L_U	1.004	0.004[a]	0.000	1.005	0.005[a]	0.000	1.004	0.004[a]	0.000
Km Secondary highways	L_S	1.003	0.003[a]	0.000	1.003	0.003[a]	0.000	1.002	0.002[a]	0.000
Distance to nearest CBD	D	1.000	0.000[a]	0.029	1.000	0.000[a]	0.002	1.000	0.000[a]	0.000
1958 [1,0]	Y_{58}				0.285	−1.256[a]	0.000	0.443	−0.815[a]	0.000
1968 [1,0]	Y_{68}	2.081	0.733[a]	0.000						
1978 [1,0]	Y_{78}				0.623	−0.473[a]	0.000	0.603	−0.506[a]	0.000
Number of obs		48,119			63,560			63,560		
Prob > chi^2		0.0000(14)			0.0000(15)			0.0000(15)		
McFadden's-R^2		0.19			0.14			0.13		
Correct predictions		Sensitivity 14.66%			Sensitivity 15.43%			Sensitivity 14.67%		
		Specificity 99.10%			Specificity 98.78%			Specificity 98.42%		
		% Correctly classified 98.23%			% Correctly classified 97.60%			% Correctly classified 96.91%		

[a]The coefficients are statistically significant at 0.10 level.

$U_C=0$; U_A=percentage of agricultural areas within each cell; U_W=percentage of water areas within each cell; L_I, L_D, L_U, and L_S=base years' highway length (km) within each cell. There are four levels of highways: L_I (interstates), L_D (divided highways), L_U (undivided highways), and L_S (secondary highways); D=distance (km) from the center of each cell to the nearest CBD; there are two CBDs, Minneapolis CBD, and St. Paul CBD; Y=dummy variable for the base year.

Results

For the three models in Table 2, the overall model is statistically significant at the 0.01 level according to the model chi-square statistic and the McFadden's R^2 ranges from 0.13 to 0.19. The cutoff point of the percentage of correct predictions (the value for determining whether an observation has a predicted positive outcome) is based on the number of the "observed=1" cells. All of the predictions are ranked from the highest to the lowest, then the top X (where X is equal to the number of the "observed=1" cells) are used as the "predicted=1" set. Since a very small proportion of the cells is "observed=1" (fewer than 2% for divided highways and fewer than 6% for secondary highways), we would expect to get a low sensitivity [Pr (predicted=1|observed=1)] and a high specificity [Pr (predicted=0|observed=0)].

The results of the predictors that test the hypotheses are summarized as follows: For population groups (P_S, P_L, P_M, and P_H), the group with the lowest population density P_S was dropped due to collinearity. The high-density group, $P_{H,}$ always has negative and significant results, which indicates that the high-density areas have a lower probability of divided highway growth than other areas. The low-density group P_L and medium density group P_M are always positive and significant, which indicates that low and medium populated areas have higher divided highway growth than other areas. These results accord with our hypothesis.

The coefficient on A is positive and significant for all the three predictions, which indicates that the new divided highways were more likely to emerge in the neighborhood of the existing corridors, and these results accord with the agglomeration hypothesis, the neighborhood of the existing corridors should be more likely to have new route development. Also for the 3-year growth prediction U_E has the highest odds ratio 2.618, which means that the neighborhood of the existing corridors are 2.618 times more likely to have divided highway development than other regions.

U_E is positive for all the three predictions and significant for the 3-year growth prediction. Since the closer the prediction the more accurate the result, we think this result generally supports the hypothesis that employment zones (including airports) have a higher likelihood of divided highway growth than other regions.

U_C is positive and significant for all the three predictions, which indicates that commercial zones are related to a higher likelihood of divided highway growth than other regions.

U_A is negative and insignificant in the 3-year and 10-year growth predictions and positive and insignificant in the 5-year growth prediction. We expected agricultural areas to be related to low growth probability since there is less traffic demand in these areas. But this can be overruled if the purpose of divided highway development is to connect urban and suburban areas and to spur economic development of the undeveloped areas. In addition, diverting from the highly urbanized areas saves construction costs.

Table 3. Logit Regression Results for Secondary Highway Growth

Dependent variable = G_S		Logit regression results								
		3_year_growth			5_year_growth			10_year_growth		
Independent variable		Odds ratio	Coef.	$P>\lvert z\rvert$	Odds ratio	Coef.	$P>\lvert z\rvert$	Odds ratio	Coef.	$P>\lvert z\rvert$
Low population area	P_L	2.064	0.725[a]	0.000	1.604	0.473[a]	0.000	1.736	0.552[a]	0.000
Medium population area	P_M	2.066	0.726[a]	0.000	1.546	0.436[a]	0.000	1.462	0.380[a]	0.000
High population area	P_H	2.592	0.952[a]	0.000	1.338	0.291[a]	0.003	1.280	0.247[a]	0.002
% Urban settlement	U_S	1.340	0.293[a]	0.080	1.634	0.491[a]	0.000	1.802	0.589[a]	0.000
Employment zones [1,0]	U_E	1.375	0.318[a]	0.000	1.281	0.247[a]	0.000	1.230	0.207[a]	0.000
Commercial areas [1,0] [[1,0][1,0]	U_C	1.269	0.238[a]	0.008	1.195	0.178[a]	0.009	1.136	0.128[a]	0.033
% Agriculture area	U_A	1.349	0.300[a]	0.050	1.428	0.356[a]	0.001	1.517	0.417[a]	0.000
% Water area	U_W	0.094	−2.360[a]	0.000	0.373	−0.987[a]	0.000	0.365	−1.007[a]	0.000
Km Interstates	L_I	1.001	0.001[a]	0.053	1.000	0.000	0.266	1.000	0.000	0.732
Km Divided highways	L_D	1.001	0.001[a]	0.090	1.000	0.000	0.126	1.000	0.000	0.627
Km Undivided highways	L_U	1.002	0.002[a]	0.000	1.001	0.001[a]	0.000	1.001	0.001[a]	0.000
Km Secondary highways	L_S	1.001	0.001[a]	0.000	1.001	0.001[a]	0.000	1.001	0.001[a]	0.000
Distance to nearest CBD	D	1.000	0.000[a]	0.001	1.000	0.000	0.106	1.000	0.000[a]	0.000
1958 [1,0]	Y_{58}				0.731	−0.314[a]	0.000	2.137	0.759[a]	0.000
1968 [1,0]	Y_{68}	2.346	0.853[a]	0.000						
1978 [1,0]	Y_{78}				0.342	−1.073	0.000	0.482	−0.729[a]	0.000
Number of obs		48,119			63,560			63,560		
Prob > chi^2		0.0000(14)			0.0000(15)			0.0000(15)		
McFadden's-R^2		0.06			0.05			0.08		
Correct predictions		Sensitivity 7.63%			Sensitivity 10.17%			Sensitivity 20.02%		
		Specificity 97.74%			Specificity 96.06%			Specificity 94.81%		
		% Correctly classified 95.53%			% Correctly classified 92.53%			% Correctly classified 90.25%		

[a]The coefficients are statistically significant at 0.10 level.

U_W is negative and significant for all the three predictions, which supports the hypothesis that water areas should be negatively related to route growth.

D is positive and significant for all the three predictions, which indicates that the farther from downtown the higher the likelihood of divided highway growth.

Secondary Highways

Hypotheses

Secondary highways carry less traffic than divided highways but more than local roads, and are composed of undivided state, county, and municipal highways. The hypotheses about the growth of secondary highways are summarized as follows.

Urban Settlements

Secondary highways serve local traffic, they are the proximate and ultimate connecting highways of urban settlements, so the growth of secondary highways should be related to the settlement areas, which include residential areas, commercial areas, industrial areas, institutions, offices, airports, and transportation infrastructure. We expect that the cells with larger settlement areas are more likely to have secondary highway growth. To test this hypothesis, the percentage of urban settlement areas within each cell is used as an independent variable U_S, and we expect this variable to be positively and significantly related to the probability of secondary highway growth.

Percentage of Water and Agricultural Areas within Each Cell

The percentage of water areas should be negatively related to the secondary highway growth probability, but we do not expect a similar relationship for agricultural areas. Agricultural areas demand product transport, and agricultural areas usually have no major (interstate or divided) highways, so secondary highways are the only commuting routes. On the one hand, the cells with a higher percentage of agricultural areas have less urban settlements and less traffic demand, which may lead to less secondary highway growth; on the other hand, however, secondary highways are more relied on in the agricultural areas for product transport and commuting service due to the lack of higher hierarchical highways, which may lead to secondary highway growth.

Model

As with divided highways, logit models are used to predict secondary highway growth based on the population distribution, land-use, and highway network data of the base years. No multicollinearity symptom has been found in the results, so we conclude that multicollinearity is not a problem in this case. Of the models tested, the following model is the best in overall model fit, and the regression results of the model are presented in Table 3

$$G_S = f(P_S, P_L, P_M, P_H, U_S, U_E, U_C, U_A, U_W, L_I, L_D, L_U, L_S, D, Y)$$

where G_S (dependent variable) = secondary highway growth, if from the base year to the predicted year there is growth in secondary highways in the observed cell, $G_S=1$, otherwise $G_S=0$;

U_S=percentage of urban settlement within each cell; here, the urban settlement include residential areas, commercial areas, industrial areas, institutions, offices, airports, and transportation infrastructure. The definitions of the other predictors are the same as those for divided highways.

Results

Of the three predictions in Table 3, the overall model is significant at the 0.01 level according to the model chi-square statistic and the McFadden's R^2 ranges from 0.05 to 0.08.

U_S is always positive and significant, which supports our hypothesis that cells with larger settlement areas are more likely to have secondary highway growth. Both employment zones (U_E) and commercial zones (U_C) have a high likelihood of secondary highway growth. Water area and their neighborhood (U_W) always have low probability of growth. The agricultural area (U_A), however, is positive and significant for all the predictions which indicates that agricultural areas are associated with a high likelihood of secondary highway growth. This result may be explained by the fact that secondary highways are more relied on in the agricultural areas for product transport and commuting service due to the lack of primary roads.

Summary and Conclusions

The study uses several decades of data from the Twin Cities Metropolitan Area to identify the land use cells that are most likely to add infrastructure (new divided and undivided highways) based on their existing state.

The results show that land-use attributes and population density level do significantly affect a cell's likelihood of adding new highway routes. New highway routes are more likely to be close to employment and commercial zones than other land-use types. Areas adjoining existing corridors have a higher likelihood of new route development. Low and medium population density areas have more highway growth than sparsely populated areas. High density is a negative for the growth of divided highways but not secondary highways. The farther from downtown, the higher the likelihood of highway growth. Cells with a higher percentage of urban settlement areas are more likely to have secondary highway growth. Agricultural areas are also associated with a high likelihood of secondary highway growth, which may be due to their high reliance on secondary highways for product transport and commuting service.

The results presented here establish that a significant portion of highway growth can be explained with easily measured variables. The success in explaining some portion of network investment suggests such decisions are not entirely political, rather investment decisions are directed based on land use factors that are in part market related. On the other hand, the limited explanatory power of the model suggests that there are other factors at work, which do include politics as well as other unmeasured variables. It is important to note that particular statistical measures (like R^2) could be arbitrarily improved by increasing cell size, without improving the real ability of the model to aid in forecasting. Future research should be directed at uncovering additional measurable factors that explain network growth decisions, and is necessary before such models are suitable for forecasting or scenario testing.

References

Garrison, W. L., and Marble, D. F. (1962). "The structure of transportation networks, transportation forecast, and prediction study progress report." *Office of Technical Services, United States Department of Commerce*, *Rep. Contract No. DA-44-177-TC-685 to the United States Army Transportation Research Command*, Washington, D.C.

Garrison, W. L., and Marble, D. F. (1965). "A prolegomenon to the forecasting of transportation development." *United States Army Aviation Material Labs Technical Rep.*, Office of Technical Services, United States Department of Commerce, Washington, D.C.

Helbing, D., Keltsch, J., and Molnár, P. (1997). "Modeling the evolution of human trail systems." *Nature (London)*, 388, 47.

Levinson, D., and Karamalaputi, R. (2003a). "Induced supply: A model of highway network expansion at the microscopic level." *J. Transp. Econ. Policy*, 37(3), 297–318.

Levinson, D., and Karamalaputi, R. (2003b). "Predicting the construction of new highway links." *J. Transp. Stat.*, 6(2/3), 81–89.

Levinson, D., and Yerra, B. (2006). "Self organization of surface transportation networks." *Transp. Sci.*, 40(2) 179–188.

Minnesota Department of Transportation. (1962, 1965, 1968, 1971, 1975, 1978, 1981, 1985, and 1990). "Minnesota official transportation maps." Minn.

Taaffe, E. J., Morrill, R. M., and Gould, P. R. (1963). "Transport expansion in underdeveloped countries: A comparative analysis." *Geogr. Rev.*, 53(4), 502–529.

Twin Cities Metropolitan Council. (1958, 1968, 1978). *Generalized land use*, Twin Cities Metropolitan Area, Minneapolis.

U.S. Census Bureau. (1960, 1970, 1980). *1960, 1970 and 1980 census tracts*, Twin Cities Metropolitan Area, Minneapolis.

Yamins, D., Rasmussen, S., and Fogel, D. (2003). "Growing urban roads." *Netw. Spatial Econ.*, 3, 69–85.

Yerra, B., and Levinson, D. (2005). "The emergence of hierarchy in transportation networks." *Ann. Reg. Sci.*, 39(3), 541–553.